UNIVERSIDAD NACIONAL DE COLOMBIA

Revista de la Facultad de Medicina 2000; 48 (2):104-110

Análisis de los datos mediante herramientas gráficas

Ricardo Sánchez Pedraza. Profesor Asociado, Centro de Epidemiología Clínica. Heidi Alexandra Cáceres. Interna Especial. Centro de Epidemiología Clínica, Facultad de Medicina, Universidad Nacional de Colombia.

Se consideran *herramientas gráficas* aquellas estructuras de organización y presentación de los datos que permiten un análisis de los mismos mediante su apreciación en el espacio. Los dos tipos de herramientas gráficas utilizados son las tablas y los gráficos propiamente dichos.

Independientemente del tipo de manejo estadístico que se vaya a dar a los datos, una etapa inicial que siempre se recomienda es la de efectuar una exploración preliminar de la información. Además de las estrategias presentadas en un escrito anterior,(1) las *herramientas gráficas* resultan de gran utilidad si consideramos las siguientes ventajas (2):

1. Permiten un mejor conocimiento de los datos, al hacer evidentes determinadas particularidades o, incluso, incoherencias que puedan llevar a revisar su calidad o a utilizar herramientas estadísticas específicas.
2. En ausencia de hipótesis (estudios exploratorios), l*as herramientas gráficas* pueden sugerir preguntas de investigación para ser analizadas en estudios posteriores.
3. En estudios diseñados para probar hipótesis *las herramientas gráficas* pueden orientar al investigador para la utilización de los métodos estadísticos más indicados.
4. En algunos casos se le debe "creer" más al gráfico que a una prueba estadística, como sucede en las pruebas de normalidad aplicadas a pocos datos.
5. Permiten transmitir la información de una manera más sencilla y clara, lo cual es de gran importancia cuando se van a presentar los resultados a un público no experto.

En general, existen dos niveles de utilización de *herramientas gráficas*: Uno en el cual se manejan como un instrumento para efectuar análisis de los datos, y otro en el cual son un medio de transmitir la información al público. En muchas circunstancias las mismas *herramientas gráficas* utilizadas para analizar datos también se utilizan para transmitir la información al público. En estos casos deben tenerse en cuenta las siguientes recomendaciones:

1. *La herramienta gráfica* debe ser lo más sencilla posible y solo debe recurrirse a ella cuando se considere que presenta los datos de una manera más clara y precisa que el texto.
2. Los datos escasos o monótonamente reiterativos no necesitan presentarse por medio de *herramientas gráficas* (3).
3. La importancia de los datos no se modifica si se añaden o no gráficos.
4. Las *herramientas gráficas* que pueden reemplazarse por pocas palabras no son necesarias.
5. La información presentada a través de una *herramienta gráfica* no tiene por qué repetirse detalladamente en el texto.
6. La información que muestra la *herramienta gráfica* debe comprenderse sin necesidad de acudir al texto (4) .
7. Los gráficos en tercera dimensión solo tienen sentido cuando la información así lo requiere. En publicaciones y presentaciones científicas debe primar lo sobrio sobre lo decorativo.

En el presente artículo se exponen los aspectos referentes a las *herramientas gráficas* como elementos imprescindibles del análisis de datos. Su utilización para publicación o difusión de resultados será objeto de una entrega posterior.

Las tablas

En general, son *herramientras gráficas* que trabajan con datos categóricos de tipo nominal u ordinal.

La forma y estructura de una tabla depende del número de variables que quiera mostrar. Cuando se maneja una sola variable se habla de distribución univariada o de frecuencia simple. Cuando la tabla cruza dos variables se está trabajando una distribución bivariada. Si la tabla maneja más de dos variables se habla de distribución

multivariada (2). Entre más variables sean introducidas en una tabla, más compleja resulta su construcción, lectura e interpretación.

Tablas para distribuciones univariadas:
La estructura típica de estas tablas es la siguiente:

Droga	Frec.	%	Acumulado
droga 1	15	25.86	25.86
droga 2	15	25.86	51.72
droga 3	12	20.69	72.41
droga 4	16	27.59	100.0
Total	58	100.00	

La estructura de la tabla consiste en cuatro columnas: la primera (Droga usada) corresponde al valor o nombre que toma la variable en esa categoría, la segunda (Frecuencia) al número de observaciones que quedan ubicadas en cada categoría, la tercera (%) al porcentaje que se ubica en las diferentes categorías, y la cuarta (Acumulado) suma consecutivamente, del primero al último, los porcentajes de la columna anterior. La información previa, presentada en forma de texto, probablemente habría sido más difícil de evaluar e interpretar. La presentación gráfica mediante tabla para distribución univariada facilita y agiliza el análisis de esta información.

Tablas para distribuciones bivariadas. Tablas cruzadas de dos variables:
Tienen la siguiente estructura:

	Severidad del cuadro			
Droga Usada	Leve	Moderado	Severo	Total
droga 1	6	4	5	15
droga 2	5	4	6	15
droga 3	3	5	4	12
droga 4	5	6	5	16
Total	19	19	20	58

Este tipo de tablas cruza dos variables categóricas. Como puede observarse, existe una fila y una columna ocupada por totales, que se denominan Totales Marginales. Estos totales marginales informan sobre la cantidad de observaciones que se encuentran en cada uno de los niveles de las variables que se están cruzando. El cruce de una fila y una columna corresponde al número de observaciones (casos, pacientes) que tienen simultáneamente dos niveles de medición de diferentes variables. En el presente ejemplo, hubo 3 pacientes que presentaban cuadro leve y recibieron la droga 3. Para complementar la información y facilitar la interpretación, se suele incluir el cálculo de porcentajes de la siguiente manera:

	Severidad del cuadro			
Droga Usada	Leve	Moderado	Severo	Total
droga 1	6 40.00	4 26.67	5 33.33	15 100.00
droga 2	5 33.33	4 26.67	6 40.00	15 100.00
droga 3	3 25.00	5 41.67	4 33.33	12 100.00
droga 4	5 31.25	6 37.50	5 31.25	16 100.00
Total	19 32.76	19 32.76	20 34.48	58 100.00

Se puede ver que el 40% de los pacientes que recibieron la droga 1 presentaron cuadros leves, o que el 37.5% de los que recibieron la droga 4 presentaron cuadros moderados, etc. Esta forma de interpretación de la tabla corresponde a una estructura de Perfil Fila. También podría generarse una estructura de Perfil Columna, como en el siguiente caso:

	Severidad del cuadro			
Droga Usada	Leve	Moderado	Severo	Total
droga 1	6 31.58	4 21.05	5 25.00	15 25.86
droga 2	5 26.32	4 21.05	6 30.00	15 25.86
droga 3	3 15.79	5 26.32	4 20.00	12 20.69
droga 4	5 126.32	6 31.58	5 25.00	16 27.59
Total	19 100.00	19 100.00	20 100.00	58 100.00

Con esta estructura es evidente que el 25% de los pacientes con cuadros severos recibieron la droga 1, que el 15.79% de los cuadros leves fueron manejados con la droga 3, etc.
Obviamente, la interpretación de la tabla mediante perfiles fila es complementaria con la interpretación mediante perfiles columna.

Tablas para distribuciones multivariadas:
Permiten el análisis de más de dos variables simultáneamente. Para incorporar más de tres variables en este tipo de estructura existen recursos como las tablas con capas que ofrecen algunos programas estadísticos (5).

La estructura más simple separa los niveles de alguna de las variables en varias tablas, tal como se muestra a continuación:

```
-> sexo=  Mujeres
                      Severidad del cuadro
Droga Usada     |  Leve   Moderado      Severo  |    Total
----------------+-------------------------------+-------
   droga 1      |    4        3            3    |       10
   droga 2      |    3        2            5    |       10
   droga 3      |    2        3     .      2    |        7
   droga 4      |    3        4            1    |        8
----------------+-------------------------------+-----------
     Total      |   12       12           11    |       35

-> sexo=  Hombres

                      Severidad del cuadro     |
Droga Usada     |  Leve   Moderado      Severo |  Total
----------------+------------------------------+----------
   droga 1      |    2       1          2      |   5
   droga 2      |    2       2          1      |   5
   droga 3      |    1       2          2      |   5
   droga 4      |    2       2          4      |   8
----------------+------------------------------+----------
     Total      |    7       7          9      |  23
```

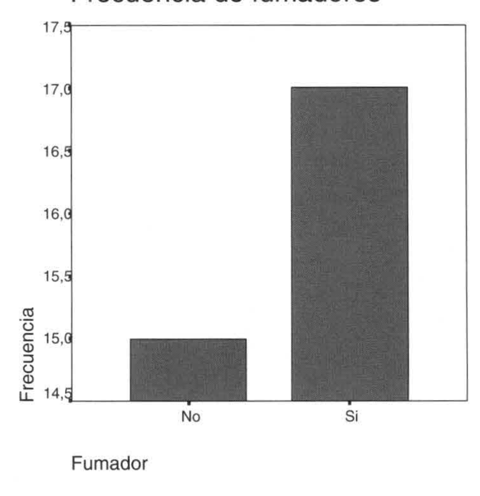
Igual que en el caso de tablas cruzadas de dos variables, puede facilitarse la interpretación con el análisis respectivo de los perfiles fila y columna.

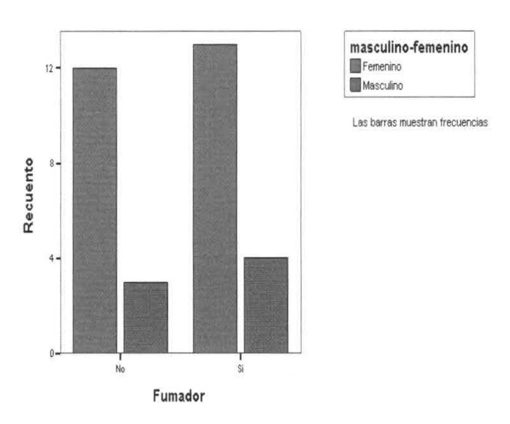
Los gráficos

Al igual que las tablas, los gráficos pueden utilizarse para análisis univariados, bivariados o multivariados. Permiten analizar más fácilmente la información de variables cuantitativas que las tablas. Las herramientas gráficas de uso más frecuente en investigación clínica se resumen en la tabla 1.

	Univariados		Bivariados		Multivariados	
	Variable cualitativa	Variable cuantitativa	Variable cualitativa	Variable cuantitativa	Variable cualitativa	Variable cuantitativa
De barras	x		x		x	
Histograma		x		x		x
De puntos	x	x				
De sectores	x		x		x	
Rama y hoja		x		x		
Caja y bigote		x		x		x
Cuartil-Cuartil		x				
De dispersión				x		x
De estrellas						x
Rostros de Chernoff					x	x

Gráfico de barras:
Representa una variable que se divide en categorías discretas, permitiendo visualizar fácilmente la distribución de las observaciones en las diferentes categorías. Cada categoría corresponde a un rectángulo con una base de tamaño constante y altura proporcional a la frecuencia en ese grupo. Generalmente los rectángulos están separados y a igual distancia uno de otro. Dentro de un análisis bivariado las barras pueden subdividirse de acuerdo con otras variables de agrupación, o ubicarse en un espacio tridimensional para efectuar análisis multivariados. Algunos ejemplos de gráficos de barras para análisis univariado y bivariado se muestran a continuación:





Histograma:
Muestra cómo se distribuyen los valores de una variable cuantitativa cuando ésta se divide en intervalos uniformes. Los rectángulos del histograma no se encuentran espaciados. Su posición corresponde a la localización del intervalo sobre el eje x y la altura indica la frecuencia dentro de cada intervalo. No se recomienda que las bases de los rectángulos tengan diferente tamaño ya que esto dificulta la interpretación de los datos (6). El histograma permite detectar valores extremos, características de simetría de la distribución o presencia de varias modas (7). El gráfico que se muestra a continuación evidencia la utilidad de esta herramienta para explorar la simetría de la distribución y la posible normalidad de una variable:

Tensión arterial sistólica

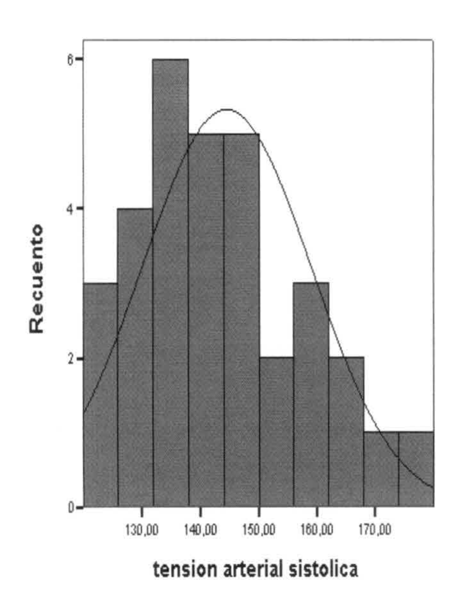


Gráfico de puntos:

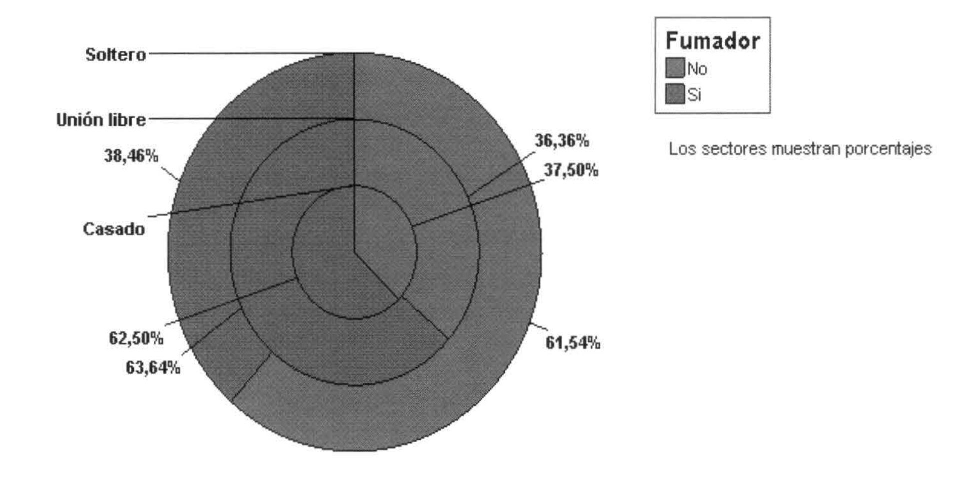
De manera similar al histograma, puede ser útil para detectar valores extremos, simetría o asimetría de los datos y localización de modas. A diferencia del histograma, el gráfico de puntos presenta los valores crudos ubicados sobre un eje continuo. El siguiente gráfico corresponde a los mismos datos del histograma presentado previamente y deja ver la presencia de un posible valor extremo:



Gráfico de sectores:

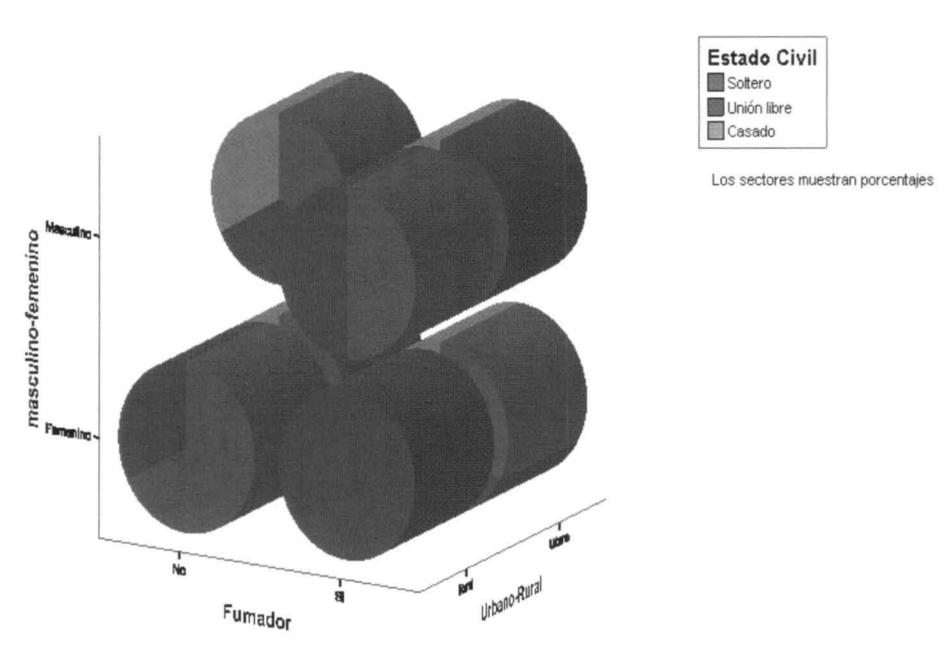
También llamados gráficos de torta o ponqués. Estos gráficos muestran la contribución de cada una de las partes a un todo, siendo el tamaño de cada una de las fracciones proporcional a su frecuencia. Permiten analizar variables cualitativas de forma uni, bi o multivariada. Una manera de utilizar estructuras de Gráfico de Sectores para análisis bivariado y multivariado sería la siguiente:

Frecuencia de consumo de cigarrillo según estado civil



Este gráfico presenta información simultánea de dos variables: Hábito de fumar y Estado civil. A medida que se introducen más variables en el procedimiento gráfico se hace más difícil el análisis, tal como se muestra en el gráfico de sectores multivariado que incorpora 3 variables: Hábito de fumar, sexo y procedencia:

Frecuencia de consumo de cigarrillo según sexo y procedencia



No se debe perder de vista que la herramienta gráfica busca facilitar en lugar de complicar el análisis de los datos.

Gráficos de rama y hoja:

Son herramientas gráficas de gran utilidad ya que tienen las mismas aplicaciones que los histogramas, pero con las siguientes ventajas adicionales:

- Preserva los valores de los datos, es decir, al observar el gráfico pueden ubicarse los diferentes valores de las observaciones.
- Es fácil de construir.
- Permite el cálculo de la mediana y el rango.

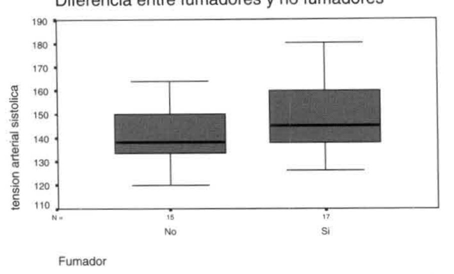
Para explicar su construcción, considérese el siguiente conjunto de datos: 10, 12, 12, 23, 24, 24, 28, 30, 30, 32, 33, 33, 34, 34, 42, 43, 43, 50, 52.

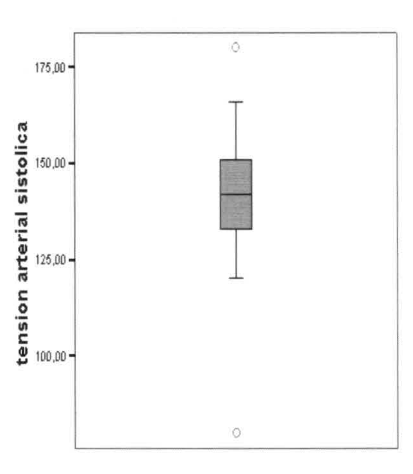
Se puede tomar como "rama" el primero de los dígitos de cada número. De tal manera puede armarse una rama con el número 1 (primer dígito de 10, 12, 12) que tendrá tres hojas (0, 2, 2), correspondientes a cada uno de los dígitos que están después del 1. Se efectúa el mismo proceso con cada una de las ramas. De esta manera el gráfico quedaría:

1 022
2 3448
3 0023344
4 233
5 02

Gráficos de caja y bigote:

Son herramientas diseñadas para explorar variables cuantitativas. La mayoría de los programas estadísticos las tienen incorporadas dentro de sus menús. La utilidad de estos gráficos radica en que permiten evaluar la simetría de la distribución, el grado de dispersión y la presencia de valores extremos. Ejemplos de este gráfico son los siguientes:

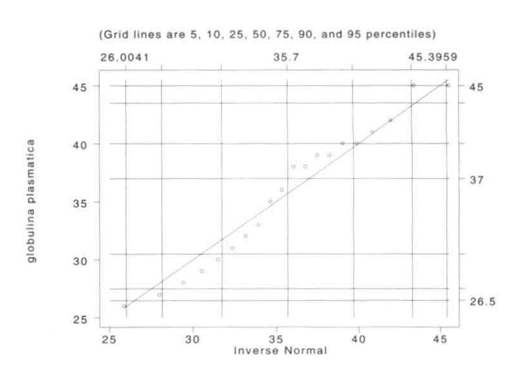


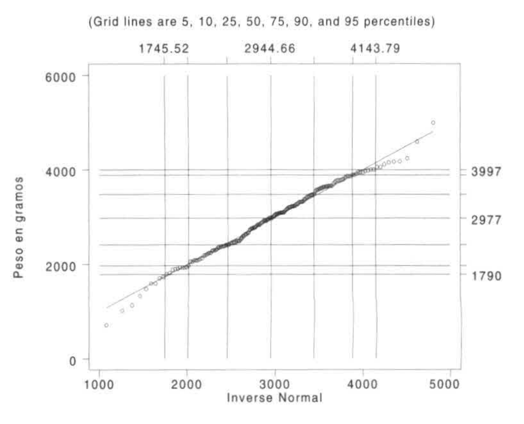


Se pueden distinguir unos cuadrados oscuros (cajas) de los cuales salen unas líneas hacia arriba y abajo (bigotes). Los límites superior e inferior de la caja están marcados por el percentil 25 y el 75 (rango intercuartílico), lo cual quiere decir que el 50% de los datos está incluido dentro de esta caja. La raya que divide a la caja en dos es la mediana. Si la distribución es simétrica, la mediana estará aproximadamente en el centro de la caja. Cada uno de los bigotes mide 1.5 rangos intercuartílicos (8). Son útiles porque los valores que queden más allá del límite del bigote pueden considerarse valores periféricos o extremos. Tales valores están representados en la gráfica por puntos.

Gráficos cuantil-cuantil:

Son herramientas utilizadas para evaluar características distribucionales de una variable, especialmente cuando estas son de tipo continuo. Es un método de aplicación fundamentalmente univariado. Estos gráficos comparan una distribución esperada, que es calculada por una función matemática determinada, con la distribución observada que es la que muestran los datos. Si la distribución de los datos se parece mucho a la distribución teórica esperada, los puntos deberán situarse muy cercanamente a una línea recta (9):

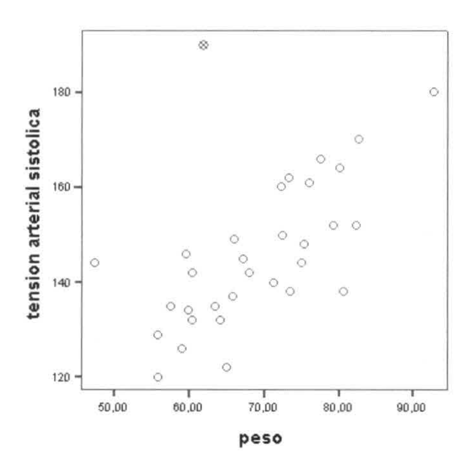




En el primer gráfico los puntos se encuentran alejados de la recta. Esto hace suponer que, en este caso, los datos no tengan distribución normal. En el segundo gráfico los puntos se encuentran muy cercanos a la recta lo que hace pensar que esta variable tiene distribución normal.

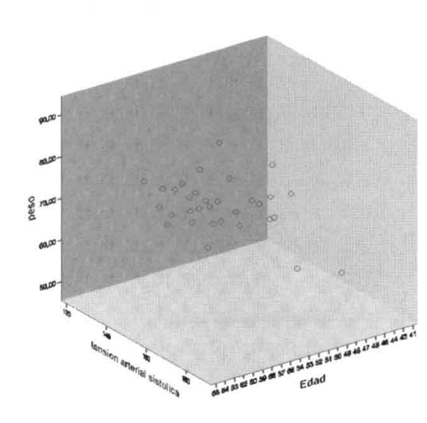
Gráficos de dispersión:
Son utilizados con mucha frecuencia para análisis bivariados donde las observaciones están representadas simultáneamente en un punto como una coordenada en un plano cartesiano. Estos gráficos son útiles en el análisis exploratorio de datos para detectar los valores extremos (10), y para explorar y describir la tendencia general de dos variables correlacionadas (11), pues muestran nubes de puntos en las cuales puede observarse fácilmente la relación entre las variables y la cercanía o lejanía de unos puntos a otros.

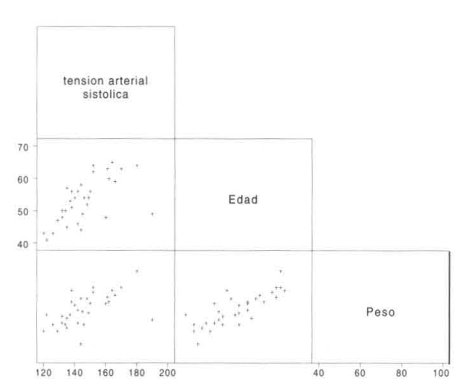
Relación entre tensión arterial y peso



El gráfico anterior muestra una clara relación entre las dos variables: a medida que aumenta el peso se incrementa la tensión arterial sistólica. Se puede ver también que existe una observación (resaltada) que se aparta de la tendencia general de los otros datos. Este tipo de gráficos suele ser el primer paso antes del desarrollo de modelos de regresión. Para el análisis multivariado con gráficos de dispersión pueden ensamblarse estructuras tridimensionales o matrices de dispersión. Un ejemplo de cada uno de estos casos se muestra a continuación:

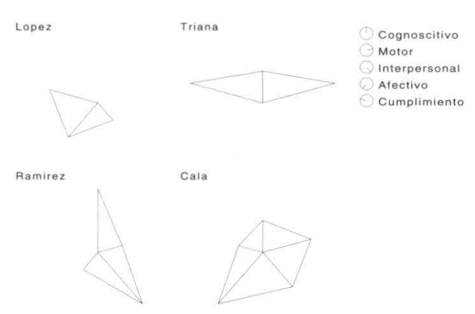






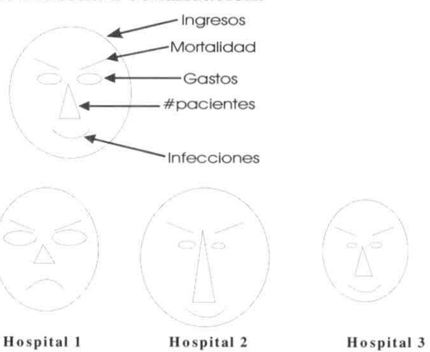
Los gráficos de tiempo son considerados una modalidad de gráfico de dispersión. Su particularidad radica en que una de las variables tiene una forma especial de continuidad. Debido a esto, al conectar los puntos con una línea se pueden apreciar tendencias o comportamientos particulares durante períodos específicos.

Gráficos de estrella:
Son gráficos que permiten la exploración visual de múltiples variables simultáneamente. No es necesario que las variables estén medidas en las mismas unidades (12). Se utilizan más que todo para comparar observaciones. Para su construcción se elaboran círculos de igual tamaño, con un número determinado de radios (igual al número de variables que se van a analizar) que tienen una longitud igual al valor de las variables. Al conectar el final de los radios se forma una estrella (10). En el ejemplo que a continuación se presenta se ilustran diferentes variables relacionadas con el cumplimiento académico de 4 estudiantes de un curso:



De acuerdo con este gráfico, el estudiante Cala tiene un perfil de desempeño más homogéneo, mientras que López y Triana solo muestran buenas habilidades en pocas áreas. Existe una variedad de este tipo de gráfico conocida como Metroglifo , en el cual los radios salen de una circunferencia pequeña y se dirigen todos en la misma dirección; su interpretación se hace de la misma manera que con los gráficos de estrella.

Rostros de Chernoff:
Teniendo en cuenta que las personas reaccionan ante las expresiones faciales, Chernoff propuso hacer representaciones multivariadas aprovechando las características del rostro. De tal manera diferentes atributos como el tamaño de la cabeza, la longitud de la nariz, la forma de la boca, etc. podrían representar a diferentes variables (10,13). Un ejemplo de Rostros de Chernoff se muestra a continuación:



Puede verse que en el Hospital 1 hay menores ingresos (cabeza más pequeña), mayores gastos (ojos más grandes), poco número de pacientes (nariz más corta) y problema de infecciones (boca de aspecto triste). Teniendo en cuenta estos parámetros puede analizarse con una rápida inspección visual la situación de los tres hospitales.

Como se ha visto, las herramientas gráficas utilizadas para el análisis de datos en investigación clínica constituyen un recurso muy valioso que ofrece grandes posibilidades. Dado el nivel de desarrollo de los programas estadísticos, se tiene al alcance la posibilidad de efectuar análisis visuales complejos y de muy buena calidad, tarea que hasta hace algún tiempo era prácticamente imposible. El análisis de los datos mediante herramientas gráficas supone, no solo una técnica, sino una actitud en la cual la ausencia de prejuicios, la curiosidad y la creatividad son elementos fundamentales (14).

Resumen:

En el presente artículo se presentan las principales herramientas gráficas para analizar datos de investigación clínica. Se muestran ejemplos de cada uno de los gráficos junto con los comentarios referentes a sus indicaciones, construcción e interpretación.

REFERENCIAS BIBLIOGRÁFICAS

1. **Quintero C, Herrera N, Sánchez R.** Conceptos básicos de Bioestadística. Revista de la Facultad de Medicina, Universidad Nacional de Colombia 1999;47(4):232-236.
2. **Bocquier P.** L'essentiel de STATA. Paris: Ritme Informatique; 1998. p 91-145.
3. **Day R.A**. Cómo escribir y publicar trabajos científicos. Washington: Organización Panamericana de la Salud; 1996. p 65-78.
4. **Squires BP.** Illustrative material: What editors and readers expect from authors. Canadian Medical Association Journal 1990; 142(5):447-449.
5. SPSS 9.0. Manual del usuario. Shouth Wacker: SPSS Inc.;1999. p173-206.
6. **Rosner B.** Fundamentals of Biostatistics. 4th ed. Belmont: Duxbury Press; 1995. p 29-30.
7. SPSS 9.0. Gráficos Interactivos. Shouth Wacker: SPSS Inc.;1999. p 103-106.
8. **Velleman PF, Hoaglin .DC.** Applications, basics and computing of exploratory data analysis.Belmont: Duxbury Press; 1981. p 65-71.
9. **Tabachnick BG, Fidell LS.** Using Multivariate Statistics. 3rd ed. New York: Harper Collins College Publishers; 1996. p 73-75.
10. **Jonson RA, Wichern DW.** Applied Multivariate Statistical analysis. 4 th ed. New Jersey: Prentice Hall; 1998. p 19-24.
11. **Rencher AC.** Methods of multivariate analysis. New York: John Wiley & Sons; 1995. p 57-59.
12. **Hamilton LC**. Statistics with STATA 5. Pacific Grove: Duxbury Press;1998. p 78-81.
13. **Gnanadesikan R.** Methods for statistical data análisis of multivariate observations. 2nd ed. New York: John Wiley & Sons; 1997. p 62-67.
14. **Ladiray D, Horber E.** Presentación de la metodología del análisis exploratorio de datos. Santafé de Bogotá: Programme PRESTA; 1999.

Nota: Tomado del libro " Estrategias de Investigación Medica Clínica "E Ardila, Sanchez, J Echeverry Eds. (en prensa).